Social Network Analysis Project

Andrea Ierardi, 960188

Code: GitHub

Dataset

- Data collected about Facebook pages (November 2017).
- These datasets represent blue verified Facebook page networks of different categories.
- Nodes represent the pages and edges are mutual likes among them.

- **Food category pages**

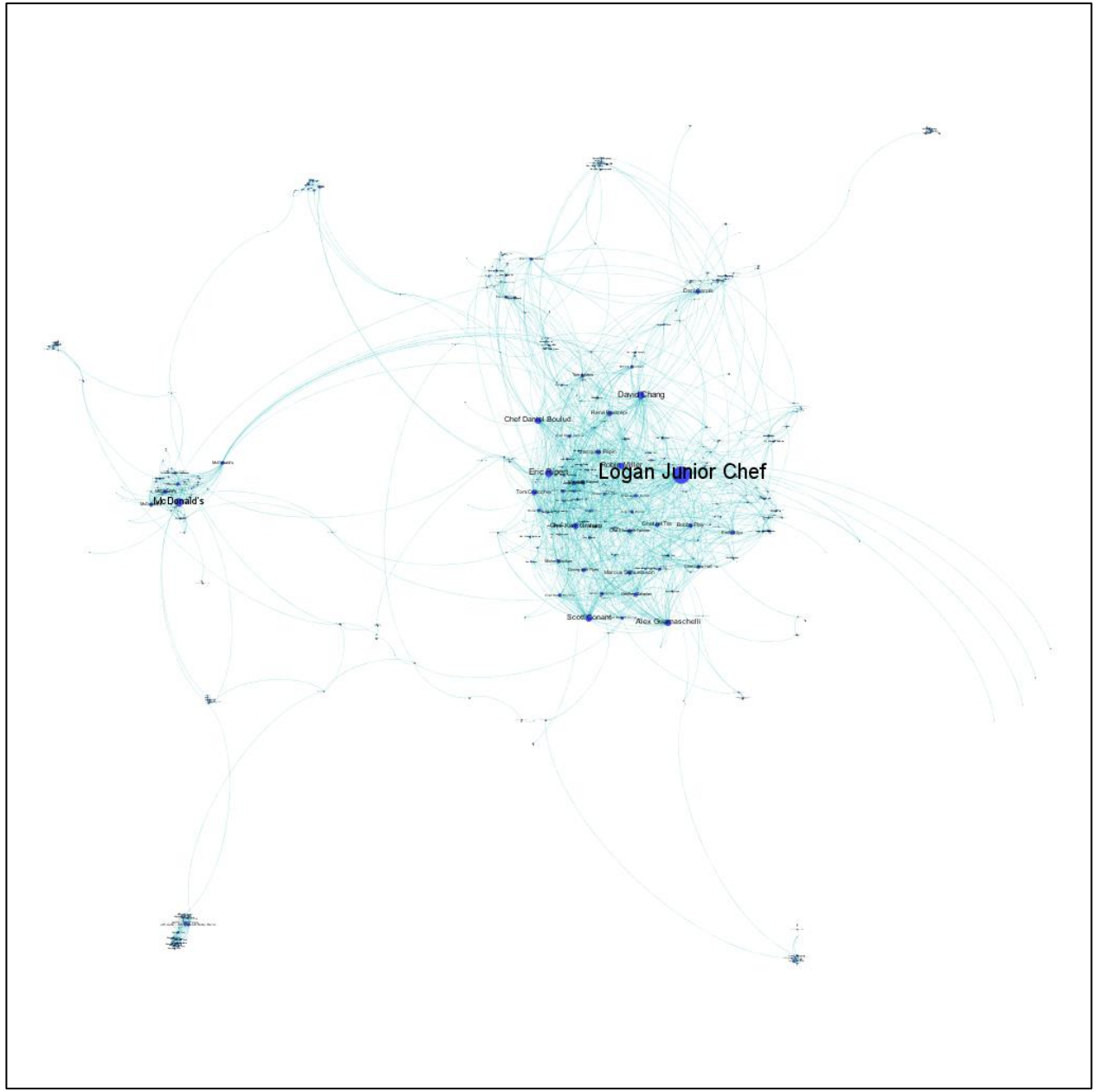
Reference: http://networkrepository.com/fb-pages-food.php

Visualisation

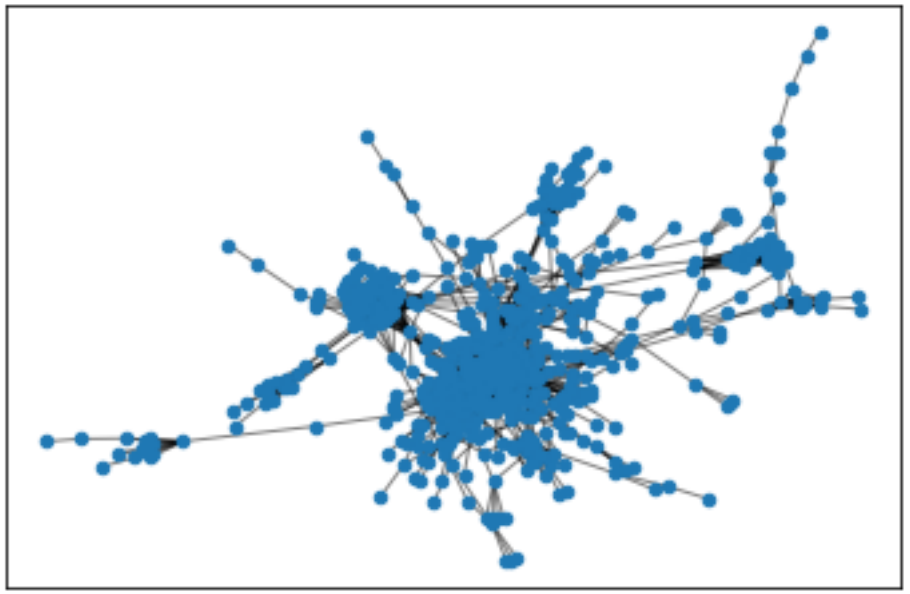
- Number of nodes: 620
- Number of links: 2102

- Undirected
- Unweighted

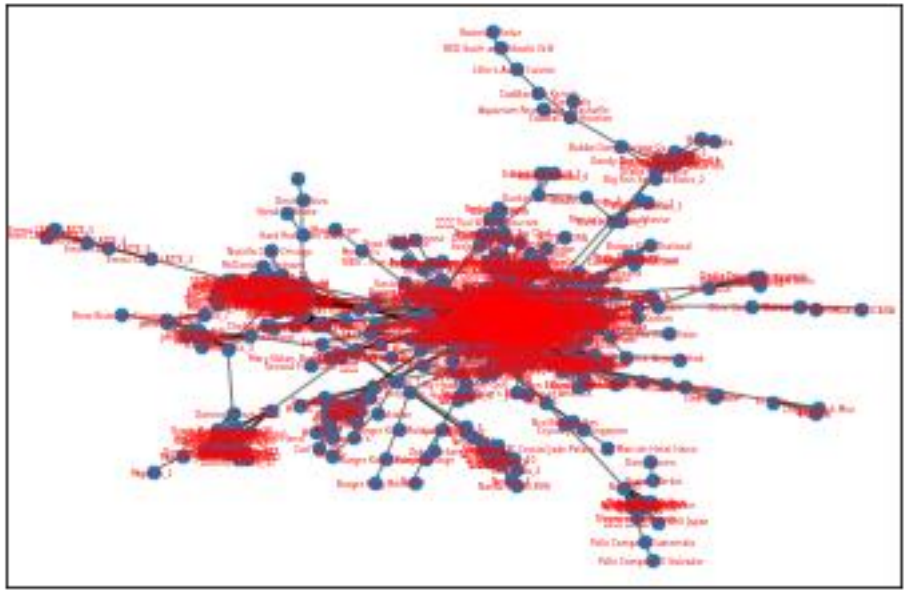
- No Isolated components



Exploration of the network



- Density: 0.01
- Standard deviation: 9.47
- Mean: 6.78
- Median: 4.0
- Min: 1
- Max: 134



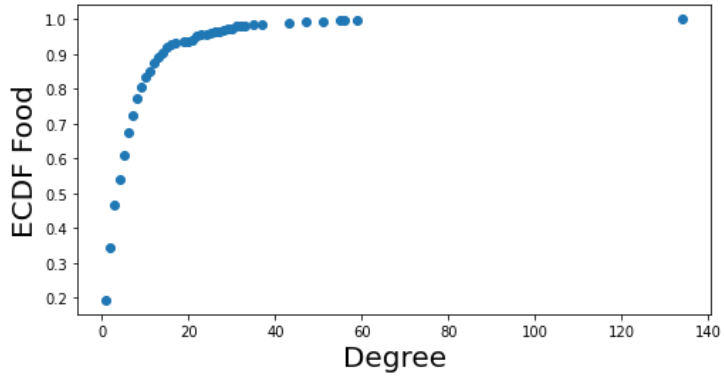
- Assortativity coefficient: - 0.028

Assortativity is a preference for a network's nodes to attach to others that are similar in some way

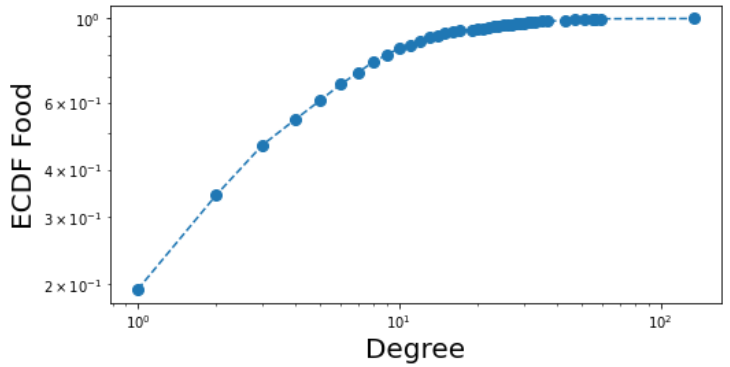
Connectivity

- No Isolated components
- Number of connected components is 1

ECDF in linear scale

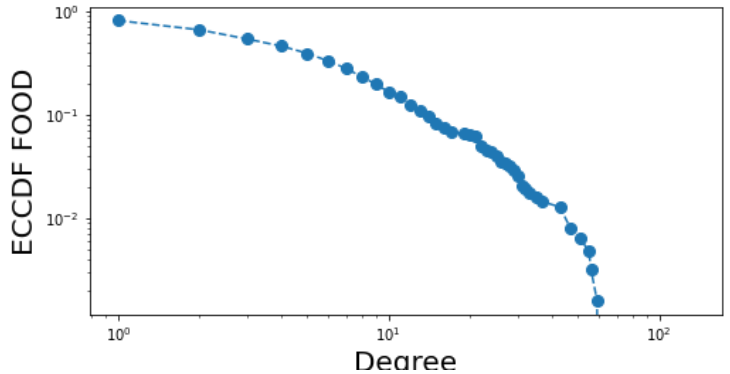


ECDF in log scale



- The number of triangle is 8805

ECCDF

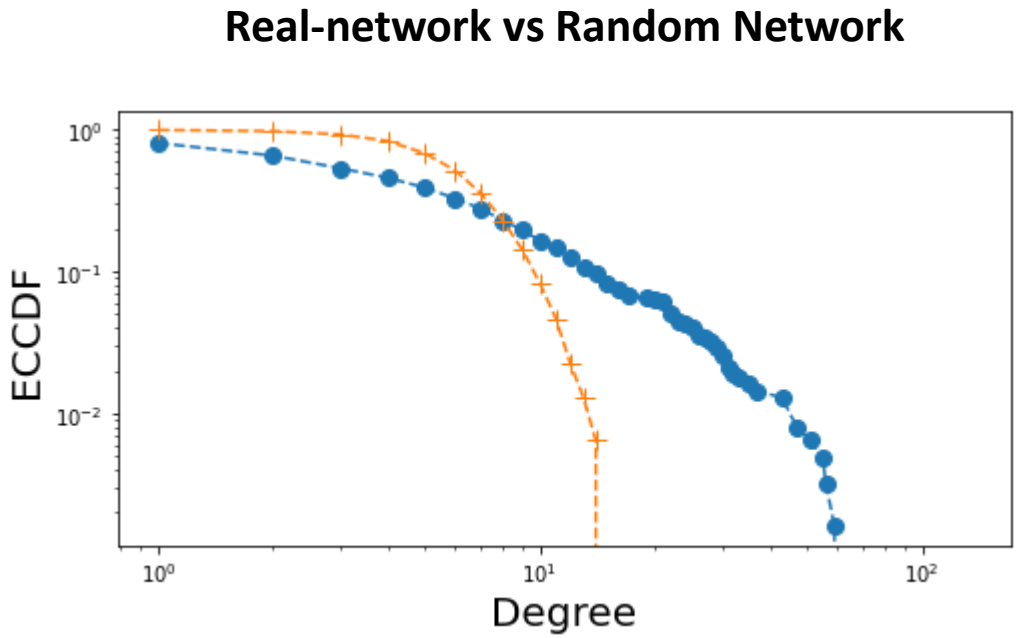


Comparison with Random Network

- This is a case of a scale-free network: follow power law Degree distribution

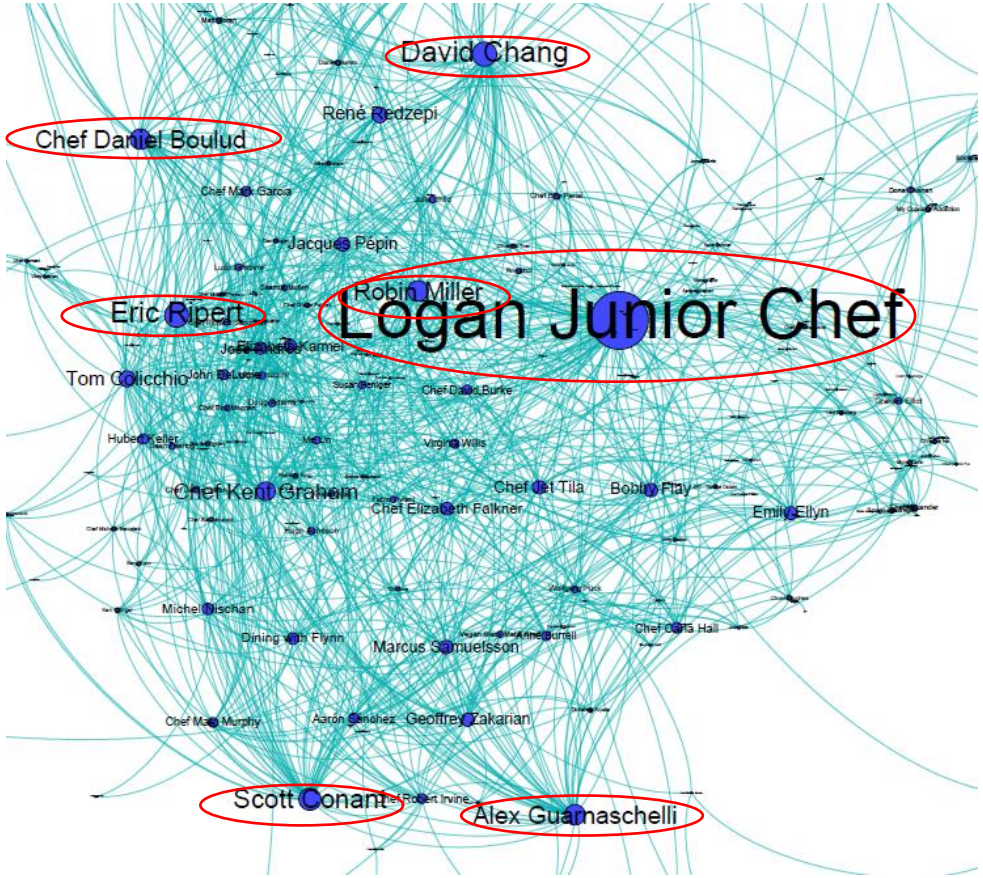
- Using a Random Network with:
  - Number of nodes: 620
  - Number of links: 2117
  - Standard deviation: 2.5
  - Mean: 6.823
  - Median: 7.0
  - Min: 1
  - Max: 15

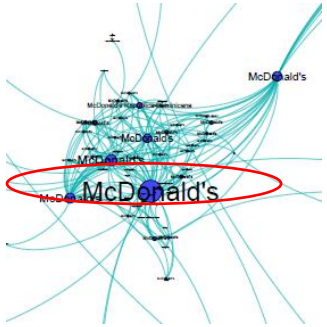
  - P = Density = Real-network density



HUBS

- 99-Percentile degree is about 47
- 8 nodes has a degree >= 47:
  - `'Logan Junior Chef','Robin Miller', 'Scott Conant', "McDonald's_6", 'Eric Ripert', 'Alex Guarnaschelli', 'David Chang', 'Chef Daniel Boulud'`





Transitivity and Clustering

- Transitivity value is : 0.223
- Average Clustering coefficient is: 0.331

Centrality

- This is based on the assumption that important nodes have many connections

Degree Centrality

```
10 most important nodes for Degree Centrality:
('Logan Junior Chef', 0.21647819063004847)
("McDonald's_6", 0.09531502423263329)
('David Chang', 0.09046849757673668)
('Eric Ripert', 0.0888529886914378)
('Scott Conant', 0.08239095315024234)
('Robin Miller', 0.07592891760904685)
('Alex Guarnaschelli', 0.07592891760904685)
('Chef Daniel Boulud', 0.07592891760904685)
('Chef Kent Graham', 0.06946688206785137)
('Tom Colicchio', 0.05977382875605816)
```

Eigenvector Centrality

```
10 most important nodes for Eigenvector Centrality:
('Logan Junior Chef', 0.3257521783526796)
('Scott Conant', 0.2212476362249716)
('Eric Ripert', 0.21990111769744997)
('David Chang', 0.2073087464245374)
('Alex Guarnaschelli', 0.20600750503290316)
('Robin Miller', 0.18493902862814815)
('Chef Kent Graham', 0.17958444309886806)
('Chef Daniel Boulud', 0.17605019181207027)
('Jacques Pépin', 0.16176590042363934)
('René Redzepi', 0.15717739346300846)
```

Betweennes Centrality

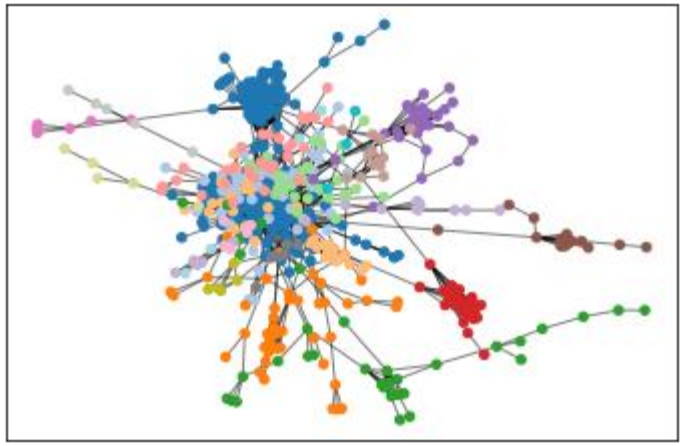
```
10 most important nodes for Betweennes Centrality:
('Logan Junior Chef', 0.3499076661737777)
("McDonald's_51", 0.1619605706800918)
("McDonald's_6", 0.14456288292404343)
('Subway', 0.09327260616363368)
('Chef Robert Irvine', 0.09141807568331686)
('foodpanda - 空腹熊貓', 0.07911166993681569)
('Dani García', 0.07229884637101391)
('達美樂披薩', 0.05838784338316884)
('Marcus Samuelsson', 0.057229302697032476)
('Chef Lorena Garcia', 0.057020665866287595)
```

Page Rank Centrality

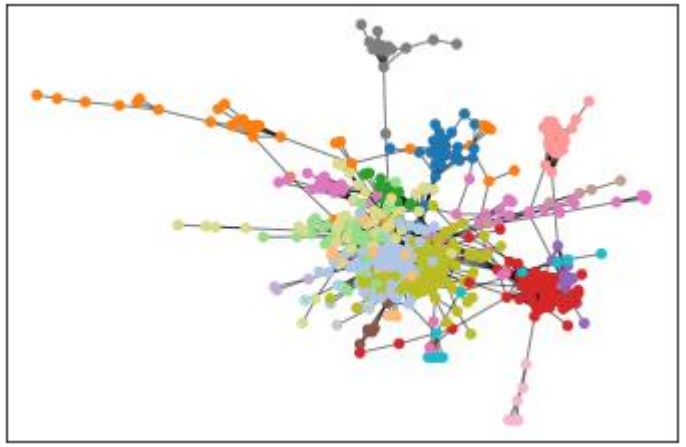
```
10 most important nodes for Page Rank:
('Logan Junior Chef', 0.025287363082514766)
("McDonald's_6", 0.014547869032715707)
('David Chang', 0.008860842463979819)
('Eric Ripert', 0.008394578695731944)
('Scott Conant', 0.00765590500196909)
('Chef Daniel Boulud', 0.007613723777791042)
('Robin Miller', 0.007566138546930535)
('Dani García', 0.007270498839968533)
("McDonald's_29", 0.007234882262368256)
('Alex Guarnaschelli', 0.007152129703475652)
```

Communities

- Using set of greedy partition we obtain 21 communities



- Using set of Louvain partition we obtain 17 communities



Coverage, modularity and performance

Greedy

```
---
Coverage 0.868696479543292 1
Modularity 0.6331957195403588
Performance 0.8936682474334254
---
```

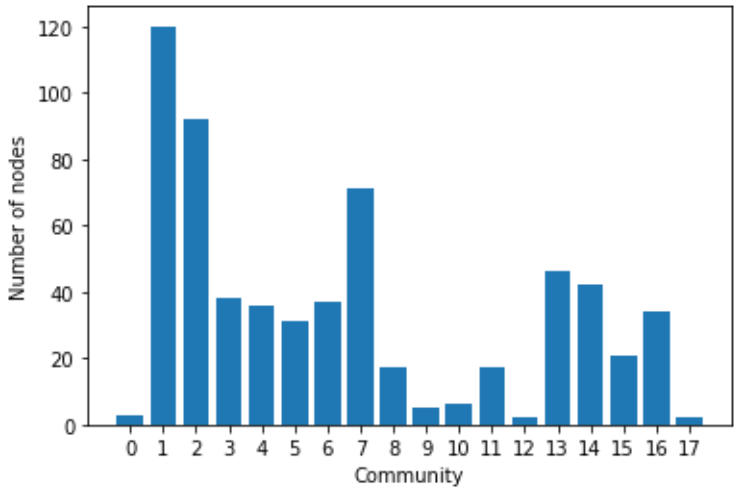
Louvain

```
---
Coverage 0.845861084681256
Modularity 0.655844621723138
Performance 0.9072906352597843
---
```

The *coverage* of a partition is the ratio of the number of intra-community edges to the total number of edges in the graph.

The *performance* of a partition is the ratio of the number of intra-community edges plus inter-community non-edges with the total number of potential edges.

Louvain Distribution



Conclusions

- No Isolated components and only one connected component
- Scale-free network
- 8 identified major HUBS
- Identified most important nodes using different Centrality methods
- Identified communities:
    - Louvain perform slightly better than Greedy in terms of performance and modularity